DOCUMENT RESUME

ED 220 488 TM 820 501

AUTHOR	Willhoft, Joseph L.; Lissitz, Robert W.
TITLE	A Methodological Study of the Torrance Tests of Creativity: Can Creativity Be Faked? Measurement and Statistics.
INSTITUTION	Maryland Univ., College Park. Coll. of Education.
PUB DATE	Mar 82
NOTE	17p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March 20-22, 1982).
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Constructed Response; *Creativity Research; Creativity Tests; Discriminant Analysis; *Experimenter Characteristics; Higher Education; Research Methodology; *Response Style (Tests); *Test Bias; Testing Problems; Test Validity
IDENTIFIERS	*Faking (Testing); *Torrance Tests of Creative Thinking

ABSTRACT

The sensitivity of the Torrance Tests of Creative Thinking with regard to an experimenter-induced response set was investigated. One hundred and ninety-eight subjects were divided into four groups. Before performing the Unusual Uses Activity (Verbal Form A), each group was given a unique set of instructions. Group I received standard instructions and acted as a control. Other treatments were varied in terms of the types of responses that were encouraged. Group II was encouraged to be "practical and reasonable"; and Group III was encouraged to list "as many ideas" as possible; and Group IV was encouraged to include all "unusual, weird, or illogical" ideas. A multivariate Dunnett test resulted in Groups III and IV being significantly different from the control group. Discriminant analysis yielded two significant functions suggesting that univariate analysis of the Torrance scales of fluency, flexibility, and originality can be misleading. The Torrance Tests were judged to be highly sensitive to experimenter bias. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220488 : 6-9

A METHODOLOGICAL STUDY OF THE TORRANCE TESTS OF CREATIVITY: CAN CREATIVITY BE FAKED?

Joseph L. Willhoft
&
Robert W. Lissitz

Annual Meeting
National Council on Measurement in Education
New York City
1982

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. W. Lissitz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MEASUREMENT
and
STATISTICS

COLLEGE OF EDUCATION
UNIVERSITY OF MARYLAND

TM 820121

ABSTRACT

The sensitivity of the Torrance Tests of Creative Thinking with regard to an experimenter-induced response set was investigated. 198 Subjects were divided into four groups. Before performing the Unusual Uses Activity (Verbal Form A), each group was given a unique set of instructions. Group I received standard instructions and acted as a Control. Other Treatments were varied in terms of the types of responses that were encouraged. Group II was encouraged to be "practical and reasonable"; Group III was encouraged to list "as many ideas" as possible; and Group IV was encouraged to include all "unusual, weird, or illogical" ideas. A multivariate Dunnett test resulted in Groups III and IV being significantly different from the Control ($p<.01$; $p<.05$). Discriminant analysis yeilded two significant functions ($p<.001$; $p<.026$), suggesting that univariate analysis of the Torrance scales of Fluency, Flexibility, and Originality can be misleading. The Torrance Tests were judged to be highly sensitive to experimenter bias.

A METHODOLOGICAL STUDY OF THE TORRANCE TESTS OF CREATIVITY: CAN CREATIVITY BE FAKED?

Joseph L. Willhoft & Robert W. Lissitz
University of Maryland

The assessment of non-cognitive traits of the individual has always been a concern in psychology and education. Interest during the 60's and 70's in affective educational programs, and current interest in indentification of the talented and gifted have resulted in the development of many instruments for measuring non-cognitive abilities. Among these measures are a number of tests of creativity (Guilford (1956), Getzels and Jackson (1962), Mednick (1962), Torrance (1966b), Wallach and Kogan (1965)). In this study we investigated the most commonly used of these tests, the Torrance Tests of Creative Thinking (Torrance, 1966).

Guilford (1956) developed a formulation of creativity in distinguishing between convergent and divergent thinking. According to his structure of the intellect, divergent processes had four dimensions: fluency, flexibility, originality, and elaboration. During the late 50's and into the mid-60's E. Paul Torrance developed a standardized test to measure creative thinking along these dimensions. In 1966 the research edition of the Torrance Tests of Creative Thinking (TTCT) was published. Since that time the TTCT has become the most widely used battery of tests of creativity. Of all the instruments contained in the Eighth Mental Measurements Yearbook (Buros, 1978), the TTCT ranked twenty-fourth in total number of references, and had the most references of all tests of creativity.

Torrance (1966b) operationalizes creativity as:

> "a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on: identifying the difficulty; seaching for solutions, making guesses, or formulating hypotheses about the difficulties; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results."

In supporting this definition, Torrance (1966b) states that he views creativity as a "generalized constellation of intellectual abilities, personality variables, and problem-solving traits." Implicit in this view is the notion that the Torrance Tests measure a fairly stable capacity of the individual.

However, there is considerable evidence that performance on tests of creativity can be influenced by conditions of testing. Elkind, Deblinger and Adler (1970) found that scores on the Wallach and Kogan Creativity Battery (1965) were affected by whether or not children were interrupted from interesting or uninteresting classroom activities in order to be tested. They identified this as a "context effect". Kirkland (1974) and Kirkland, Kilpatrick and Barker (1976) replicated these findings using the Torrance Tests. In a similar study, Boersma and O'Bryan (1968) found that subjects given the TTCT in the same room where they had been given an intelligence test on the previous day scored significantly lower than subjects given the TTCT in a dissimilar setting.

The findings of motivational effects on TTCT performance are more equivocal. While Halpin and Halpin (1973) were able to improve TTCT scores under conditions of increased motivation, Cox, Nash and Ash (1976) found no effects when TTCT performance was linked to rewards.

In addition to antecedant conditions and motivational effects, some researchers have focused on the instructions that are given to test takers. Most of this research has been limited to comparing standard instructions with anxiety-producing instructions. Bamber, Jose and Boice (1975) found that "differential instructions (as a test, as a game, without instructions)...did not affect creativity" among undergraduate subjects, although such effects had been found with children (Bamber, 1973). Similarly, Trentham (1972), using only the Originality scores of the TTCT, found no difference between subjects given regular instructions and subjects given anxiety-producing (i.e., achievement-oriented) instructions.

The present study was designed to see if TTCT creativity

scores could be affected by subjects' perceptions of the criteria for successful performance. We hypothesized that by altering the standard instructions (to emphasize different types of responses as being acceptable) we would be able to affect creativity scores. The analogy we used was to ask ourselves what effect a similar experiment might have on a measure of academic achievement. For example, could one expect a difference in, say, mathematics achievement scores as a result of subjects being told that they should demonstrate extra mathematics ability? We felt that no such differences could be anticipated on a valid achievement test. However, as Crockenburg (1972) pointed out in her review of tests of creativity, "(w)hile we cannot readily dismiss the Torrance Test, neither is it reasonable to conclude that it is a valid measure of the creative process". If our hypothesis were supported, we felt this would be further evidence against the validity of the TTCT, and would support the notion that a major component of creativity scores is the response-set under which test takers operate.

PROCEDURE

198 subjects participated in this study. All subjects were graduate students in the College of Education at the University of Maryland and were enrolled in either a beginning or an intermediate statistics class. The nine sections of these two classes were randomly assigned to one of the four treatments. This resulted in the following cell distribution: Treatment I--two sections, 49 subjects; Treatment II--two sections, 49 subjects; Treatment III--two sections, 46 subjects; Treatment IV--three sections, 54 subjects.

The instrument used in this study was Activity Five (Unusual Uses:Cardboard Boxes) of Verbal Form A of the TTCT. In this activity subjects are given 10 minutes to list as many new and unusual uses for cardboard boxes as they can think of. Performance in this activity is scored on three scales: Fluency, Flexibility, and Originality. Fluency is defined as: "the test

taker's ability to produce a large number of ideas with words." Flexibility is defined as: "a person's ability to produce a variety of kinds of ideas, to shift from one approach to another, or to use a variety of strategies." Originality is defined as: "the subject's ability to produce ideas that are away from the obvious, commonplace, banal, or established." (Torrance, 1966)

Under typical administration procedures of the TTCT, prior to opening their test booklets subjects are read a set of instructions about how to perform in general. For this experiment, the four treatments consisted of three variants on these instructions plus the standard set. Inserted in each test booklet was a copy of the treatment instructions. In addition to these instructions, all test booklets had a set of directions printed at the top of the page that were specific to the Unusual Uses Activity. After opening their booklets, subjects read along as the experimenter read both sets (treatment instructions and activity instructions) aloud. The "activity instructions" that were common to all treatments were:

> "Most people throw their empty cardboard boxes away, but they have thousands of interesting and unusual uses. In the spaces below and on the next page, list as many of these interesting and unusual uses as you can think of. Do not limit yourself to any one size of box. You may use as many boxes as you like. Do not limit yourself to the uses you have seen or heard about; think about as many new uses as you can."

The unique instructions given to Treatment I were the standard instructions typically given to test takers as a guide to performance on the TTCT (Torrance, 1966a). This treatment was viewed as a Control group. This group's instructions were:

> "The activity in this booklet will give you a chance to use your imagination in thinking up ideas and putting them into words. There are no "right" or "wrong" answers like there are in most things that we do. We want you to see how many ideas you can think of and we think you will find this fun. Try to think of interesting, unusual, and clever ideas--something that no one else will think of."

The above set of instructions contains four sentences. An analysis of the content of these sentences was used to design the three variant instruction sets for Treatments II, III, and IV. The first sentence was viewed as a general introduction to the task and was essentially left intact in the three variants. The

second sentence was viewed as encouraging test takers to be unrestricted in their responses. We interpretted the third sentence as placing importance on the number of responses and as encouraging enjoyment. We viewed the fourth sentence as somewhat restrictive, by virtue of its emphasis on the uniqueness of responses.

Treatment II, the first variant, was designed to be considerably more restrictive than the standard set. These instructions were:

> "The activity in this booklet will give you a chance to think of ideas and put them into words. We urge you to concern yourself with the practicality and reasonableness of your ideas."

Here, the first sentence deleted the reference to one's imagination. Also deleted were references both to the number of ideas and to enjoying the task. The "uniqueness" restriction in the standard instructions was altered to be even more restrictive by asking test takers to only attend to "practical and reasonable" ideas.

Treatment III, the second variant, was designed to emphasize the "number of ideas" aspect of the standard instructions. This set was:

> "The activity in this booklet will give you a chance to use your imagination in thinking up ideas and putting them into words. We want you to see how many ideas you can think of for this activity."

The last variant, Treatment IV, was designed to be an expanded version of the standard instructions that would more strongly urge people to include unique ideas. This set read as follows:

> "The activity in this booklet will give you a chance to use your imagination in thinking up ideas and putting them into words. We have found it helpful to encourage people to use their imagination while doing this activity. Remember, this is a test of creativity and that means there are no "right" or "wrong" answers. Feel free to include ALL your ideas, no matter how impractical or silly they may seem to you.
>
> "Our experience has been that whether or not you ARE creative is not as important as whether or not you THINK YOU ARE creative. Thus, we strongly urge you to adopt a posture of being creative, to think of yourself as a creative person, to imagine how a creative person might respond. Remember, we are interestted in answers that make you look as creative as possible, so try try to think of interesting, unusual, and clever ideas--something that no one else will think of.

"Remember to relax, to think of yourself as a creative person, to record all your ideas no matter how unusual, weird, or illogical they may seem, and HAVE FUN!"

Our expectation was that Treatment II would be the lowest scoring group on the three scales of Fluency, Flexibility, and Originality; that Treatment III would be the highest on Fluency and Flexibility; that Treatment IV would be the highest on Originality; and that Treatment I (Control) would show median performance on all three scales.

All test situations were administered by the same experimenter. The standard time limit for this activity was adhered to. Test booklets were numbered only, and subjects were told that their performance would remain anonymous and would in no way affect their grade in the course.

Test booklets were scored by the Torrance Tests Scoring Service. The Scoring Service was informed neither of the nature of the study nor of the coding scheme for numbering the test booklets.

RESULTS

Raw score means and standard deviations for the four groups on each of the three scales are presented in Table I. Because the TTCT Norms-Technical Manual (Torrance, 1966b) "recommend(s) that users base their interpretations on the three...verbal scores" (p. 72), and because our hypotheses about the relative performance of the four groups were based on these scales, three

insert Tables I & II about here

univariate F-tests were performed to test for differences among the four groups. These results are shown in Table II. Inspection of Table I indicates that our hypotheses were generally borne out. The major discrepency was that Treatment I (Control) performed lowest on all three scales. Aside from this, however, Treatment II was the lowest of the three variant groups on each of the three scales; Treatment III was the highest of all groups on Fluency and Flexibility; and Treatment IV was the highest on Originality. The ANOVA F-tests indicated that significant group differences occurred on all the scales.

However, the anomalous performance of Treatment I was puzzling. Why, for example, did this group score lower on Fluency than the more restricted Treatment II? An attempt to better understand this phenomenon and the high within-cells correlations between the scales (see Table III) led us to a multivariate analysis of

insert Table III about here

the data. The considerable overlapping of the traits being measured, as evidenced by the high intercorrelations among the scales, suggested that there were not three distinct dimensions to TTCT performance. Hotelling's T-squared statistic was computed for contrasts comparing the three variant treatments with the Control Treatment. Critical values were obtained using a multivariate extension of the univariate Dunnett test, which simultaneously controls for Type I error and is, consequently, more powerful than the F-test usually reported with T-squared statistics (Higazi and Dayton, 1982). The III vs. I contrast resulted in a T-square=16.83 ($p<.01$); for IV vs. I, T-square=13.44 ($p<.05$); and for II vs. I, T-square=7.28 ($p>.05$). Given that there were significant differences between two of the three variant treatments and the Control, the question of the nature of these differences arose. A discriminant analysis was performed and resulted in two significant functions ($p<.001$ and $p<.026$) and one non-significant function ($p<.625$). The correlations between the two significant functions and the original TTCT scales are shown in Table IV. These correlations

insert Table IV & Figure 1 about here

suggest that the first function is primarily composed of Flexibility and Fluency with a sizable contribution from Originality. Recalling how the TTCT scales are defined, this function appears to measure how many different ideas one has and the number of ideas one has. The high correlation of Originality with both Fluency and Flexibility (Table III) suggests that that portion of Originality that correlates with Fluency and Flexibility is also part of this first function. The second

function is most highly correlated with Originality, and nearly uncorrelated with both Flexibility and Fluency. Remembering how Originality is defined, this function seems to distinguish between those who have common and banal ideas and those who have unique and unusual ideas, quite apart from the number of, or differences between, one's responses. Plots of the group means in the 2-space of the discriminant variables are presented in Figure 1. Here, it can be seen that the major difference between Treatment III and Control is determined by the first function. Recalling that group III was urged to come up with as many ideas as they could, this difference makes sense. Treatment IV differs from Control on both the first function (though not by as much as does Treatment III) and the second function. This was the group that was urged to record all their ideas no matter how illogical they seemed. It appears that these instructions resulted in performance that yielded more ideas and, at the same time, more unique ideas than did the Control instructions. Treatment II does not appear to be very distinct from the Control in this 2-space, and this is supported by our finding no significant difference here. This group's variant instructions were a recommendation to think of ideas that were practical and reasonable. It can be argued that this set of restrictions is not unlike the standard instructions which place restriction on responses ("Try to think of interesting, unusual, and clever ideas--something that no one else will think of").

In fact, the number of restrictions that were put upon appropriate responses may have been a critical factor in this study. Group III was given the least restrictive instructions, resulting in the highest (overall) scores. Treatment IV, while encouraged to include all ideas, was restricted by directions to "think of interesting, unusual and clever ideas". This group scored the next most creative of the four. The standard instructions (Treatment I) had the same restrictions as Treatment IV, yet did not have the encouragement that the latter group had. Lack of a significant difference between Treatments I and II supports the observation that both of these low-scoring groups

were similarly restricted.

DISCUSSION

We are aware that this study has some limitations and that its results should be viewed in light of them. First, the entire TTCT battery of tests was not used; we only used Activity 5 of the Verbal Form A. While one might suspect that the results would generalize to the entire battery, further study using all the activities is called for. Second, the Control group for this study was not a true control in that its treatment deviated from standard administration practices for the TTCT. Subjects in this study were given copies of the test instructions while standard practice is to have the instructions read to them. Again, one might suspect that having the experimenter read the instructions would allow for more expectancy effects than providing subjects with their own instructions would. Finally, intact groups were randomly assigned and degrees of freedom were based upon an assumption of random assignment of subjects to groups. Analysis of test scores covering course material (quizzes, midterms, finals, projects) revealed no academic differences between the groups, and we can think of no compelling reason to support the notion that the groups had pre-existing differences on creativity dimensions. Keeping these points in mind, we suggest the following conclusions.

First, two of the three variant treatment groups differed significantly from the Control group. This suggests that performance on the TTCT is extrememly sensitive to the response set under which test takers operate. It would be patently absurd to claim that our instructions to the groups actually altered their basic ability to be creative. Our data suggest that what test takers think is being asked for will have a profound effect on how they will perform. Secondly, this problem is confounded by the fact that most applications of the TTCT are not "blind" studies, and that in very many cases the test administrator is well aware of the results that are desired. Thus, strong effects

due both to the experimenter expectancies and to subjects' response sets cannot be ruled out.

Third, there do not appear to be three distinct creativity traits measured by the TTCT. Our data suggest that there are only two dimensions to creativity. Cicirelli (1964) and Long and Henderson (1964) have also reported high correlations between the TTCT scales of Fluency, Flexibility and Originality (ranging from .60 to .80). We agree with Thorndike's (1972) observation that "the evidence of consistently different meaning for the fluency, flexibility, and originality scores is almost vanishingly small", and feel that a univariate approach to data analysis can be misleading. We recommend that researchers be very cautious when interpretting TTCT scale scores as being measures of the traits as defined in the test manual.

Our first point seems to be the most critical one. The degree to which test takers feel restricted or encouraged on the responses may well have a critical effect on their performance. That our modest treatments were able to induce significant differences in creativity performance is quite striking. One way to interpret these differences is to view our treatments as "mini" training programs on how to be creative. Our data suggest that reseachers interested in evaluating the effectiveness of creativity training cannot use a base-line of "no change" as a null criterion. Our findings indicate that the "no difference" condition for the Torrance Tests may have an inherent variability of at least one standard deviation (see Table I). It is known that experimental subjects may react to an experimenter's expectations (Orne, 1969; Rosenthal, 1969). Our finding of the extreme sensitivity of the TTCT to experimenter-induced response sets leads us to conclude that even under conditions of rigorous control, studies using the Torrance Tests should be viewed with extreme caution. Not only may the results be a methodological artifact, but the interpretation of creativity as a "real" internal characteristic of a person is subject to suspicion.

REFERENCES

Bamber, R.T. Play, interest, domestication, and creativity. Unpublished doctoral dissertation. University of Missouri, 1973.

Bamber, R.T., Jose, P.E., & Boice, R. Creativity as affected by differential reinforcements and test instructions. Bulletin of the Psychonomic Society, 1975, 6, 361-363.

Boersma, F.J., & O'Bryan, K. An investigation of the relationship between creativity and intelligence under two conditions of testing. Journal of Personality, 1968, 36, 341-348.

Buros, O.K. (ed.). The eighth mental measurements yearbook, vol. I. Highland Park, New Jersey: Gryphon Press, 1978.

Cicirelli, U.G. The relationship between measures of creativity and academic achievement: Interaction and threshold effects. Unpublished doctoral dissertation, University of Michigan, 1964.

Cox, R.S., Nash, W.R., & Ash, M.J. Instructions for three levels of reward and creativity test scores of college students. Psychological Reports, 1976, 38, 411-414.

Crockenburg, S.B. Creativity tests: A boon or a boondoggle for education? Review of Educational Research, 1972, 42, 27-45.

Elkind, D., Deblinger, J., & Adler, D. Motivation and creativity: The context effect. American Educational Research Journal, 1970, 7, 351-357.

Getzels, J.W., & Jackson, P.W. Creativity and intelligence. New York: Wiley & Sons, 1962.

Guilford, J.P. The structure of the intellect. Psychological Bulletin, 1956, 53, 267-293.

Halpin, G., & Halpin, G. The effect of motivation on creative thinking abilities. Journal of Creative Behavior, 1973, 7, 51-53.

Higazi, S.M.F., & Dayton, C.M. Comparing several experimental groups with a control in the multivariate case. Paper presented at the Annual Meeting of the Biometric Society, San Antonio, Texas, 1982.

Kirkland, J. On boosting divergent thinking scores. California Journal of Educational Research, 1974, 25, 69-72.

Kirkland, J., Kilpatrick, A., & Barker, W. Sex difference in boosting divergent thinking scores by the context effect. Psychological Reports, 1976, 38, 430.

Long, B.H., & Henderson, E.H. Opinion formation and creativity in elementary children. Unpublished research report, University of Delaware, 1964.

Mednick, S.A. The associative basis of the creative process. Psychological Review, 1962, 69, 220-232.

Orne, M.T. Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R.L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969.

Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R.L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969.

Thorndike, R.L. Review of Torrance tests of creativity. In O.K. Buros (ed.), The seventh mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1972.

Torrance, E.P. Torrance tests of creative thinking: Directions manual and scoring guide. Princeton, New Jersey: Personnel Press, 1966.

Torrance, E.P. Torrance tests of creative thinking: Norms-technical manual. Princeton, New Jersey: Personnel Press, 1966.

Trentham, L.L. Anxiety levels, distraction conditions, and instruction types in a creativity testing situation. Unpublished doctoral dissertation, Indiana University, 1972.

Wallach, M.A., & Kogan, N. Modes of thinking in young children. New York: Holt, Rinehart & Winston, 1965.

	Fluency	Flexibility	Originality
Fn. 1:	-.859	-.906	-.678
Fn. 2:	-.156	-.345	-.662

Table IV: Correlations between discriminant functions and TTCT scales.

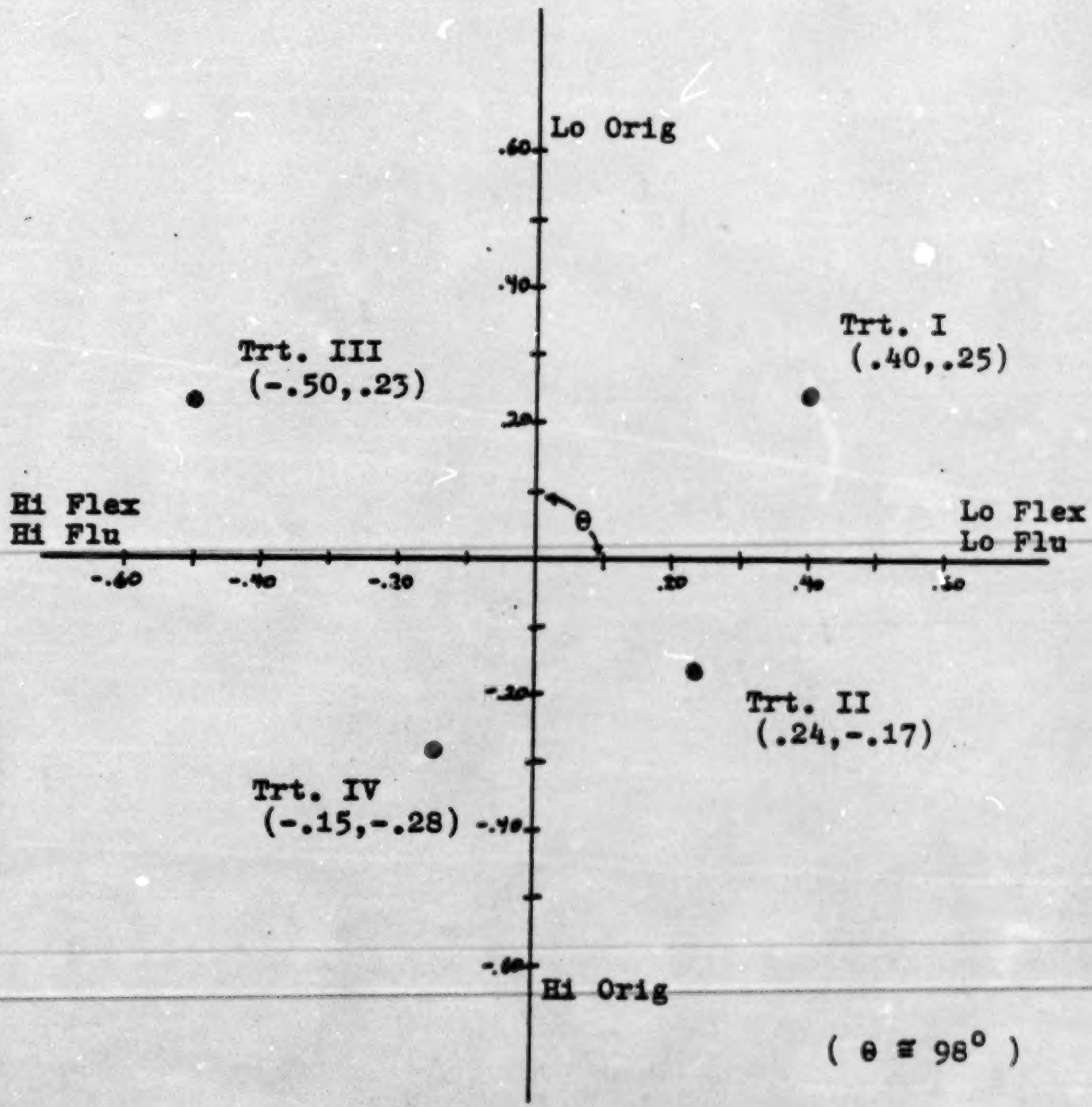

Figure 1. Plots of group means on 1st and 2nd discriminant functions.

		FLUENCY	FLEXIBILITY	ORIGINALITY
Treatment I	MEAN:	22.53	11.63	23.76
n = 49	S.D.:	8.06	3.46	10.90
Treatment II	MEAN:	24.00	12.65	28.20
n = 49	S.D.:	7.28	2.78	12.18
Treatment III	MEAN:	29.22	14.35	31.09
n = 46	S.D.:	9.78	3.42	11.05
Treatment IV	MEAN:	27.54	13.87	32.67
n = 54	S.D.:	10.64	3.48	14.47

Table I: Means and standard deviations for the 4 groups on each of the 3 TTCT scales.

TTCT Scale	Mean Square Error	F (3,194)	p less than
Fluency	62.270	5.602	.001
Flexibility	10.933	6.672	.001
Originality	151.625	5.068	.002

Table II: Univariate F-tests for group differences on each of the 3 scales.

	Flu	Flex	Orig
Fluency	1.000		
Flexibility	.687	1.000	
Originality	.838	.749	1.000

Table III: Within-cells correlations among the scales.